[This Page Intentionally Left Blank]

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS OF *IN VITRO* CYTOTOXICITY TEST METHODS AND THE ABILITY OF THESE TEST METHODS TO PREDICT ACUTE SYSTEMIC TOXICITY

In vitro cytotoxicity test methods based on NRU have been evaluated for a number of uses. This section reviews studies relevant to:

- the prediction of acute rodent systemic toxicity using *in vitro* NRU cytotoxicity test methods
- the use of *in vitro* cytotoxicity test methods to predict starting doses for acute systemic toxicity tests, and
- the use of *in vitro* NRU cytotoxicity test methods to predict other *in vivo* endpoints.

Section 9.1 discusses *in vitro* studies that evaluated cytotoxicity using NRU for correlation with acute systemic toxicity in rodents and with other *in vivo* endpoints. Also reviewed are studies that have evaluated the use of *in vitro* cytotoxicity results to reduce animal use in acute toxicity testing. **Section 9.2** reviews independent evaluations of the use of *in vitro* cytotoxicity methods to determine starting doses for acute systemic toxicity assays and a validated NRU test method similar to that used in the current study. The conclusions of these reports will be compared to the conclusions reached in this study where possible. **Section 9.3** reviews studies that have used the *Guidance Document* approach (ICCVAM 2001b), which establishes the current test method performance standard.

9.1 Relevant Studies

9.1.1 Correlation of *In Vitro* NRU Cytotoxicity Results with Rodent Lethality

This section reviews *in vitro* cytotoxicity studies that have used NRU methods to predict rodent lethality. *Italics* identify chemicals tested in the reviewed studies that were also tested in the NICEATM/ECVAM validation study reviewed in this BRD.

Peloux et al. (1992)

Using several different *in vitro* cytotoxicity test methods with primary rat hepatocytes, Peloux et al. (1992) determined the correlation with rat/mouse intraperitoneal (ip) or intravenous (iv) LD_{50} values for the 25 chemicals tested. The *in vitro* cytotoxicity test methods, which used a 20-hour chemical exposure duration, assessed the following endpoints: NRU; total protein content, lactate dehydrogenase (LDH) release, tetrazolium salt MTT reduction. [NOTE: MTT is metabolized by mitochondrial succinate dehydrogenase of proliferating cells to yield a purple formazan reaction product.] The IC_{50} values obtained using the four endpoints were highly correlated (r = 0.973-0.999) to one another. For the IC_{50}-LD_{50} regressions, Peloux et al. (1992) used the lowest reported LD_{50} value published for rat or mouse studies that administered the test substances acutely using the ip or iv routes. The regressions used units of ln μg/mL for the IC_{50} and ln mg/kg for the LD_{50}. The IC_{50} values obtained using NRU had the highest correlation coefficient, r = 0.877, to the to rat/mouse ip/iv LD_{50} values. The total protein assay yielded r = 0.872, the MTT reduction assay yielded r = 0.808, and the LDH release assay yielded r = 0.789.

Fautrel et al. (1993)

Six laboratories tested the cytotoxicity of 31 chemicals in primary rat hepatocytes using a 24-hour exposure followed by measuring NRU. The investigators performed linear regression analyses for the prediction of rat iv, ip, and oral LD_{50} values by the NRU IC_{50} values. The regressions by the various *in vivo* administration routes did not use the same chemicals since LD_{50} values for all the routes were not available for all the tested chemicals. Oral , iv, and ip LD_{50} values were available for 27, 24, and 18 chemicals, respectively. IC_{50} values were obtained for 15, 14, and 11 of the chemicals, respectively. The units used for correlation were ln μg/mL for the IC_{50} and ln mg/kg for the LD_{50}. While the regression for the iv data was statistically significant (r = 0.88, n=11), the ip (r=0.48, n=14) and oral regressions (r=0.17, n=15) were not. The fact that the parenteral LD_{50} values correspond more closely with *in vitro* cytotoxicity data than do the oral LD_{50} was thought to be due to the fact that there are fewer kinetic variables (i.e., absorption, distribution, etc.) to consider for iv administration. The authors concluded that the hepatocyte cultures were useful in screening chemical classes with high bioavailability.

Roguet et al. (1993)

Roguet et al. (1993) tested the cytotoxicity of 28 MEIC chemicals in primary rat hepatocytes exposed to the chemicals for 21 hours, followed by measuring NRU. A correlation of the NRU IC_{50} values to LD_{50} values obtained from the unpublished data of B Ekwall et al. yielded a statistically significant linear correlation ($p < 0.001$) with $r = 0.80$. [NOTE: The LD_{50} values subsequently published by Ekwall et al. (1998) were from the 1997 edition of RTECS®.] The correlation used molar units for the *in vivo* and *in vitro* data. Roguet et al. (1993) reported that the toxicities of thioridazine, malathion, and *copper sulfate* were overestimated and the toxicity of *potassium cyanide* was underestimated, but their criterion for over/under estimation was not provided. The toxicity of *potassium cyanide* was also underpredicted (see **Appendix L-2**) when using the Registry of Cytotoxicity (RC) rat only weight regression (i.e., $\log LD_{50} = 0.372 \log IC_{50} + 2.024$) prediction of GHS toxicity categories by the NICEATM/ECVAM 3T3 and NHK NRU test methods. The RC is a database of acute oral LD_{50} values for rats and mice obtained from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for chemicals with known molecular weights (Halle 1998).

Rasmussen (1999)

Twenty MEIC chemicals were tested for cytotoxicity in 3T3 cells using NRU with and without the addition of Arochlor-induced rat liver microsomes (S9 mix). The chemical exposure duration was 24 hours. Similar to the present validation study, Rasmussen (1999) was unable to attain cytotoxicity with *xylene*, although it was dissolved in ethanol instead of DMSO. In the presence of S9, the cytotoxicities of malathion, 2,4-dichlorophenoxyacetic acid, *propranolol*, thioridazine, *lithium* sulfate, *copper sulfate*, and *thallium sulfate* were significantly decreased ($p<0.05$) while the cytotoxicities of *1,1,1-trichloroethane*, *phenol, nicotine*, and *paraquat* were significantly increased ($p<0.05$).

The toxicities of *nicotine, thallium sulfate,* and *paraquat* were also underpredicted in the NICEATM/ECVAM validation study (see **Appendix L-2**) when using the RC rat only

weight regression (i.e., log LD_{50} = 0.372 log IC_{50} + 2.024) prediction of GHS toxicity categories by 3T3 and NHK NRU test methods.

Although both IC_{20} and IC_{50} values were determined in the Rasmussen (1999) study, only the IC_{20} values were used for correlations with rat acute oral LD_{50} values from RTECS®. Even though the units of the LD_{50} values were not reported, the correlations are assumed to be in molar units since the IC_{20} and IC_{50} values were reported in μM units. Significant linear correlations ($p < 0.001$) for IC_{20} and LD_{50} values were obtained with and without microsomes. The correlation was slightly higher with microsomal activation ($r = 0.72$ vs. 0.68 for oral and 0.82 vs. 0.78 for ip).

Although the presence of S9 increased the cytotoxicity of some chemicals, it decreased the toxicity of others, and yielded only a small improvement in the correlation to *in vivo* data.

Creppy et al. 2004

Creppy et al. (2004) used a 48-hour NRU assay to determine the cytotoxicity of ochratoxin A (OTA) and fumonisin B1 (FB1) on C6 glioma (rat brain), Caco-2 (human intestinal), and Vero (green monkey kidney) cells. The IC_{50} determined in the NRU assay was used in the RC regression to predict the acute oral LD_{50}. The predicted LD_{50} using the C6 glioma cells was similar to mouse LD_{50} values (data generated from four *in vivo* studies), but the LD_{50} values predicted by the other cell lines were about 50 times greater than that predicted by the C6 glioma cells. The authors found the relative insensitivity of the Vero cells surprising since OTA was known to be a kidney toxin. There were no LD_{50} values with which to compare the predicted LD_{50} of FB1.

9.1.2 Use of Cytotoxicity Data to Reduce the Use of Animals in Acute Toxicity Testing

Halle et al. (1997): Animal Savings Using Cytotoxicity Data with the ATC

This study predicted the animal savings that would be produced by using IC_{50} data from cytotoxicity tests in the RC regression to determine a starting dose for ATC testing. No cytotoxicity testing was performed for this study. The authors used the IC_{50x} data from the RC and the RC regression to predict the LD_{50} for the 347 RC chemicals. At the time of the

Hall et al. (1997) study, the ATC (1996 version from OECD) was designed to classify chemicals using the three classes of acute oral toxicity and an unclassified group defined by the acute oral toxicity classification system of the European Union (EU) (see **Table 9-1**). Thus, the fixed doses for the ATC testing were 25, 200, and 2000 mg/kg.

Table 9-1 EU[1] Classes of Acute Oral Toxicity

Category	LD_{50} (mg/kg)
1	$LD_{50} \leq 25$
2	$25 < LD_{50} \leq 200$
3	$200 < LD_{50} \leq 2000$
Unclassified	$LD_{50} > 2000$

[1]Anon (1993)

Halle et al. (1997) used the RC predicted LD_{50} for the 347 RC chemicals as a starting point to estimate the number of ATC dose steps (and animals required) that would be needed to classify the chemicals in the same EU category associated with *in vivo* LD_{50} (i.e., oral rat or mouse values from RTECS®). The method required the simulated ATC testing for each chemical to start at the nearest fixed ATC dose to the LD_{50} predicted by the RC. The outcome of the simulated testing of three animals per fixed dose was determined by the *in vivo* LD_{50}. If the test dose was lower than the *in vivo* LD_{50}, animals were assumed to live and, if the test dose was higher than the LD_{50}, the animals were assumed to die. Testing of the chemical would proceed with higher (when the animals lived) or lower fixed doses (when the animals died) until the chemical was placed into the EU toxicity category indicated by the *in vivo* rodent oral LD_{50}.

The method of Halle et al. (1997) can be illustrated with digoxin, which has an *in vivo* rodent LD_{50} of 18 mg/kg (from RTECS®) and an RC predicted LD_{50} of 414 mg/kg. Simulated ATC testing would start at the nearest fixed dose, 200 mg/kg, to the RC predicted LD_{50} of 414 mg/kg. The three animals tested would die, and three more animals would be tested at 25 mg/kg. The animals tested at 25 mg/kg would die and digoxin would be classified in category 1 for $LD_{50} \leq 25$ mg/kg. Thus, classification of digoxin required six animals.

Using such simulations of ATC testing for the 347 RC substances, Halle et al. (1997) estimated a total of 2139 test animals would be used:

- 328 substances would require testing with two doses with three test animals each
- 19 substances would require testing with three doses with three animals each

Halle et al. cited (from Schlede et al. 1995) that the average number of animals required to classify chemicals using the ATC method was 9.11. Using this average, ATC testing of the 347 RC chemicals would require 3161 animals. Thus, there would be a 32% reduction in the number of test animals used (compared to the average) when the RC LD_{50} prediction was used in conjunction with the 1996 version of the ATC method (Halle et al. 1997).

Depending on the regression evaluated, the average animal savings for the ATC predicted in the NICEATM/ECVAM validation study at dose-response slopes of 2 and 8.3 were 8.0 – 14.8% (0.85-1.56 animals) for the 3T3 NRU and 8.9 -13.5% (0.97-1.43 animals) for the NHK NRU for the 72 reference substances tested (see **Section 10.3**). This is quite a bit lower than the average savings of 32% calculated by Halle et al. (1997). However, there were a number of differences between the evaluation performed by the Halle et al. (1997) and the NICEATM/ECVAM study that contribute to the difference in calculated animal savings:

- the NICEATM/ECVAM study used *in vitro* cytotoxicity data to estimate starting doses (using several regressions based on the RC data)
- the chemicals tested in the NICEATM/ECVAM study were different from the RC chemicals (i.e., the 58 RC chemicals tested had a regression significantly different from the RC regression [see **Section 6.1.2**])
- the NICEATM/ECVAM study used computer simulations of ATC testing, which incorporated assumptions about mortality distributions, to determine animals used whereas Halle et al. (1997) used simplified assumptions (i.e., animals live when test dose is less than LD_{50} and die when test dose is greater than LD_{50})
- the NICEATM/ECVAM study determined animal savings by comparing animal use with starting doses determined by the *in vitro* data to animals used at the

default starting dose of 300 mg/kg. Halle et al. (1997) used the average animal use for the ATC as a comparison to animal use with simulated testing..

- the NICEATM/ECVAM study used the GHS acute toxicity categories for classification whereas Halle et al. (1997) used the EU toxicity classification scheme, which had fewer toxicity categories (i.e., accuracy of category prediction by any method would be higher with fewer categories).

Spielmann et al. (1999): Animal Savings Using Cytotoxicity Data with the UDP

Spielmann et al. (1999) recommended an *in vitro* cytotoxicity procedure for reducing the number of animals used in acute toxicity tests. The procedure used *in vitro* cytotoxicity data as a range finding test for the *in vivo* toxicity test.

The authors identified nine chemicals in common when comparing the RC database to an evaluation of acute toxicity methods by Lipnick et al. (1995). Spielmann et al. (1999) compared the LD_{50} values from Lipnick et al. (1995) to LD_{50} predictions calculated when using the RC IC_{50} values in the RC regression formula. For seven of the nine chemicals, the LD_{50} prediction was within an order of magnitude of the conventional LD_{50} (OECD 1987) used in Lipnick et al. (1995). Spielmann et al. (1999) concluded that the RC provides an adequate prediction of LD_{50} and that cytotoxicity data could be used to predict starting doses for the UDP. If an IC_{50} is available for a particular chemical, the authors recommend using the IC_{50}, with the RC regression, to calculate a starting dose (i.e., estimated LD_{50}) for the UDP, FDP, or ATC method.

If no IC_{50} is available for a particular chemical, Spielmann et al. (1997) recommended determining cytotoxicity using a standard cell line and specific endpoint of cytotoxicity (e.g., NRU, total protein, MTT reduction, etc.). To show that the *in vitro* cytotoxicity test methods provide results that are consistent with the RC, Spielmann et al. (1999) recommended testing 10-20 RC chemicals. The IC_{50} data are then used to calculate a new regression, which is then compared to the RC regression. If the new regression fits into the acceptance interval (± log 5 of the fitted regression line) of the RC regression line, the RC regression is used to predict starting doses for the UDP. If the new regression is parallel to the RC regression, but outside

the ± log 5 acceptance interval, Spielmann et al. (1999) recommended using the new regression line for the prediction of the starting dose.

Spielmann et al. (1999) contends that the RC provides a sufficient prediction of LD_{50} values from IC_{50} values for chemicals that do not require metabolic activation and are not usually toxic (i.e., $LD_{50} > 200$ mg/kg), such as industrial chemicals. The authors acknowledge that the fit of chemicals with $LD_{50} < 200$ mg/kg to the RC regression is not good and attribute the poor fit of these chemicals to the fact that they require metabolic activation for toxicity. They indicated that the prediction of starting doses using cytotoxicity data can be applied to the UDP and ATC methods, but not to the FDP since dosing is not sequential (this contradicted a claim made earlier in the paper that the approach could be used with the FDP). They did not estimate the number of animals that might be saved with this approach, but they did recommend that the approach be validated experimentally using several established cell lines with a limited number of representative chemicals from the RC.

EPA (2004): U.S. EPA HPV Challenge Program Submission

PPG Industries, Inc. is the manufacturer of Propanoic acid, 2-hydroxy-, compd. with 3-[2-(dimethylamino)ethyl] 1-(2-ethylhexyl) (4-methyl-1,3-phenylene)bis[carbamate] (1:1) [CAS No. 68227-46-3] and is the sponsor of this compound for the EPA HPV Chemical Challenge Program. The compound is an isolated intermediate and subsequently is used to produce a resin component of paint products. PPG provided the following data on the compound in their submission (http://www.**epa**.gov/chemrtk/prop2hyd/c13863rt3.pdf) to the EPA: physical-chemical, environmental fate and pathway, ecotoxicity, and toxicological. The acute mammalian toxicological data were generated using *in vitro* and *in vivo* test methods.

An *in vitro* NRU cytotoxicity test with BALB/c 3T3 cells was conducted to estimate a starting dose for the *in vivo* acute oral toxicity study using the UDP (OECD 2001a) (see **Appendix M** for the OECD UDP test guideline). Use of *in vitro* methods was intended to minimize the number of animals used for *in vivo* testing. The estimated LD_{50} of the compound determined by the NRU assay was 489 mg/kg. The starting dose for the UDP study was set at 175 mg/kg, the first default dose below the estimated LD_{50} value. The

starting dose of 175 mg/kg is also the default starting dose, which is used when no information (on which to base a starting dose) is available . A total of fifteen female rats received the compound at 175, 550, or 2000 mg/kg. Five of nine rats treated at 2000 mg/kg died prematurely on Days 2 and 3. At 2000 mg/kg, 2/4 surviving animals had lost up to 25% of their Day 1 body weights by Day 15. The LD_{50} for the compound was estimated to be 2000 mg/kg with a 95% confidence interval of 1123-5700 mg/kg. Thus, the *in vitro* NRU cytotoxicity test method overpredicted the toxicity of the compound by estimating a lower LD_{50} value than that determined in the acute oral toxicity UDP study. The report authors felt that a greater than predicted number of animals was used for UDP testing since the LD_{50} estimated by the 3T3 NRU assay, 489 mg/kg, and, consequently, the starting dose, was much lower than the *in vivo* LD_{50} of 2000 mg/kg. However, since the UDP started with the default starting dose of 175 mg/kg, the claim that more animals were used is unfounded, since animal use with the default starting dose is the baseline with which animal use should be compared.

9.1.3 Other Evaluations of 3T3 or NHK NRU Test Methods

This section briefly reviews studies that have evaluated NRU methods for purposes other than the prediction of starting doses for acute oral systemic toxicity assays. NRU test methods using either 3T3 or NHK cells have been evaluated for use as alternatives to the Draize eye irritation test and to predict acute lethality in humans. Except for the 3T3 NRU phototoxicity assay, NRU methods have neither been scientifically validated by an independent review for any of these purposes nor accepted for regulatory use. The use of the 3T3 NRU test method to determine phototoxic potential is addressed in **Section 9.2** since it has been validated.

Based on the method of Borenfreund and Puerner (1985), the *in vitro* NRU test method protocols evaluated in the reviewed studies are similar to those evaluated in the current study. The major difference is that most use a 24-hour chemical exposure duration for the 3T3 assay, while the current 3T3 validation study used a 48-hour exposure duration. The major difference between the NHK NRU test method protocols used in these studies and the

protocol used in the NICEATM/ECVAM study is the change of medium with test chemical application used in the validation study presented in this BRD.

Draize eye irritation

Triglia et al. (1989)

Four laboratories collaborated in an interlaboratory validation study to test the NHK NRU assay from Clonetics® Corporation. The evaluation included intra- and inter-laboratory reproducibility and the ability to predict *in vivo* ocular irritancy. Each laboratory tested 11 surfactant-based test agents and compared the IC_{50} values to available *in vivo* Draize ocular irritancy scores.

The authors determined the following performance characteristics when comparing the *in vitro* and *in vivo* data:

- specificity (percentage of non irritants detected) = 93%
- sensitivity (percentage of true irritants detected) = 80%
- predictive values (probability that an unknown agent will be properly classified)
  - positive predictive value = 90%
  - negative predictive value = 87%

The authors concluded there was excellent correlation among the laboratories and good correlation between the *in vitro* NR_{50} values (concentration at 50% reduction of NRU compared to controls) and the Draize data (Spearman Rank correlation coefficients between *in vivo* and *in vitro* data for the laboratories ranged from 0.67-0.76). The authors also concluded, however, that the NRU assay could not replace the Draize test but may be an effective screening tool for use in a battery of *in vitro* alternatives.

Sina et al. (1995)

Sina et al (1995) evaluated the NHK NRU test method along with six other *in vitro* test methods to evaluate whether they could be used as complimentary tests in a battery approach. The NRU data correlated poorly with Draize scores for the 33 pharmaceutical intermediates that were tested. The Spearman correlation coefficient for the NR_{50} and

maximum average Draize score (MAS) was -0.10 and the Pearson correlation coefficient was -0.04.

Brantom et al. (1997)

This study examined the potential of 10 alternative methods to predict the eye irritation potential of cosmetic ingredients. Four laboratories tested 55 coded substances (23 ingredients and 32 formulations) with the 3T3 NRU test method and used the resulting IC_{50} to predict modified maximum average scores (MMAS) for the Draize test.

An endpoint in µg/mL was generated for each test by interpolation from a plot of percentage cell survival versus the test substance concentration. A prediction model (PM) was developed from data of 30 single ingredients (29 surfactants and one chemical not classified by the authors) to equate the IC_{50} value to an MMAS.

The interlaboratory CV for the NR_{50} values was 37.3 ± 29.8% (7.5 ± 6.8 log transformed). No mean IC_{50} value for a single laboratory differed by an order of magnitude from the mean of all the laboratories for each chemical, which the authors interpreted as "no significant outliers". Correlations of NRU predicted MMAS scores with *in vivo* MMAS scores yielded Pearson's r = 0.25 – 0.32 (for the four laboratories).

Although the authors concluded the reproducibility was good, the data did not accurately predict the MMAS (i.e., low r values for *in vitro*/*in vivo* correlations; underpredicted irritants, overpredicted non-irritants). However, the authors concluded that the 3T3 NRU test method had wide applicability to test 51/55 coded substances according to the limitations in the prediction model (four substances outside of the 95% confidence intervals), but that it was not a stand-alone replacement for the Draize test across the entire irritation scale. None of the substances tested were identified by the authors.

Harbell et al. (1997)

This paper reported the results of the evaluation of 12 *in vitro* cytotoxicity assays to predict ocular irritation. Data were voluntarily submitted to the US Interagency Regulatory

Alternatives Group (IRAG), composed of members from CPSC, EPA, and FDA. The NHK NRU test method was one of the tests evaluated by six laboratories testing surfactants and surfactant-containing formulations (the 3T3 NRU was not tested). Two laboratories submitted results for the same test substances, but the other four laboratories submitted data for various sets of chemicals and formulations.

The correlation of results from the two laboratories testing the same substances was r = 0.99. Correlations between the NR_{50} data and *in vivo* maximum average score (MAS) ranged from -0.92 to -0.54. The IRAG concluded that the assay was suitable as a screening and adjunct assay to assess eye irritation over the range of toxicities found in personal care and household products. IRAG recommends that its use be limited to water-soluble materials. Although the method was evaluated for surfactants, IRAG recommended that the evaluation continue for its performance in predicting eye irritation for various product classes (e.g., fabric softeners, shampoos). IRAG also recommended that physical form be considered since toxicity of the solution (*in vitro*) does not necessarily predict toxicity of the solid (*in vivo*).

Predicting human lethal blood concentrations

Seibert et al. (1992)

The aim of this single laboratory study was to evaluate various aspects of cellular toxicity in four *in vitro* test systems for relevance and reliability to acute systemic toxicity, in particular, human lethal blood concentrations. The 3T3 NRU test method was one of four methods evaluated with 10 MEIC chemicals.

The authors stated that final conclusions on the relevance of the *in vitro* systems could not be determined when compared to the *in vivo* data. The variations in lethal blood concentrations are unknown and make it difficult to define limits for over/underprediction of *in vivo* toxicity using experimental models. In addition, the ability of *in vitro* toxicity to predict *in vivo* toxicity may strongly depend on toxicokinetic factors.

9.2 Independent Scientific Reviews

This section **(9.2)** covers independent scientific reviews of the use of *in vitro* cytotoxicity methods for the prediction of acute oral toxicity and reduction of animal use. The conclusions of these reports are compared to the conclusions of the current study. Also discussed is the 3T3 NRU phototoxicity test method that has been validated by ECVAM.

9.2.1 Use of *In Vitro* Cytotoxicity Data for Estimation of Starting Doses for Acute Oral Toxicity Testing

ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the ATC

Participants at Workshop 2000 examined the influence of starting dose on animal usage for the ATC method (ICCVAM 2001a, section 2.2.3, pp.12-14). No testing was performed at the Workshop. The participants made inferences from the 1996 version of the ATC method that was based on the EU hazard (i.e., toxicity) classification system in **Table 9-1**. The fixed doses for testing were 25, 200, and 2000 mg/kg. Normally, classification of a substance requires testing three animals in two to four dosing steps (i.e., six to 12 animals). With increasing distance between the true toxicity class and the starting dose, the number of dosing steps increases. They estimated that one to three dosing steps could be avoided if the optimum starting dose could be predicted by *in vitro* cytotoxicity (i.e., three to nine animals saved).

The savings of one to three dosing steps was predicted under ideal conditions. The Workshop 2000 report (ICCVAM 2001a) provides a biometrical analysis at a dose-mortality slope of 2 by W. Diener that shows that the largest animal savings occur for chemicals with very high and very low toxicity. Three animals are needed to classify a chemical in the < 25 mg/kg class if the true LD_{50} is 1 mg/kg and 25 mg/kg is the starting dose, but six animals are needed if the test starts from the default starting dose of 200 mg/kg (i.e., animal savings = 33%). For a chemical with a true LD_{50} of 10000 mg/kg, 11.3 animals on average are needed using the default starting dose, but only 7.7 animals are needed at the 2000 mg/kg starting dose (i.e., animal savings = 31%). For chemicals with a true LD_{50} of 2000 mg/kg, no animals are saved by starting at the 2000 mg/kg dose (compared to starting at the default starting dose of 200 mg/kg).

Although these analyses were performed assuming the 1996 ATC method used starting doses of 25, 200, 2000 mg/kg, Workshop 2000 participants expected that animal savings that would be produced by improving the starting dose would not be significantly different for the current ATC method that uses GHS doses of 5, 50, 300, and 2000 mg/kg (or up to 5000 mg/kg) (OECD 2001c; see **Appendix M** for the current ATC test guideline).

Beyond presenting the biometrical analysis by W. Diener, Workshop 2000 participants did not predict the animal savings when *in vitro* cytotoxicity methods are used to estimate starting doses for the ATC.

The NICEATM/ECVAM study yielded a pattern of animal savings for the ATC that was similar to those discussed at the 2000 Workshop (i.e., animal savings were greater for chemicals with lower or higher LD_{50} than the default starting dose; see **Section 10.3**). Depending on the regression evaluated, the average animal savings (for the 72 reference substances tested) predicted by the NICEATM/ECVAM validation study at a dose-response slope of 2 was:

- 12.8-17.1 % (1.22-1.63 animals) for the 3T3 NRU and 7.6-13.0% (0.72-1.23 animals) for the NHK NRU for chemicals in the $LD_{50} \leq 5$ mg/kg category
- 12.1-16.6 % (1.45-1.98 animals) for the 3T3 NRU and 18.9-23.9% (2.26-2.86 animals) for the NHK NRU for chemicals in the $5 < LD_{50} \leq 50$ mg/kg category
- 3.6-4.3 % (0.39-0.47 animals) for the 3T3 NRU and 2.1-2.8% (0.23-0.30 animals) for the NHK NRU for chemicals in the $50 < LD_{50} \leq 300$ mg/kg category
- -2.8- -0.2% (-0.24 - -0.02 animals) for the 3T3 NRU and -1%-0.8% (-0.10-0.02 animals) for the NHK NRU for chemicals in the $300 < LD_{50} \leq 2000$ mg/kg category
- 1.4-14% (0.16–1.67 animals) for the 3T3 NRU and 3.4%-11.0% (0.38-1.23 animals) for the NHK NRU for chemicals in the $2000 < LD_{50} \leq 5000$ mg/kg category
- 16.2-31.1% (1.92-3.70 animals) for the 3T3 NRU and 14.2-29.2% (1.69-3.47 animals) for the NHK NRU for chemicals with $LD_{50} > 5000$ mg/kg

The major differences between the evaluation reviewed by the Workshop 2000 participants and the NICEATM/ECVAM study were:

- the NICEATM/ECVAM study used *in vitro* cytotoxicity data to estimate starting doses (using several regressions based on the RC data), whereas the Workshop 2000 participants used the fixed ATC doses as starting doses
- the NICEATM/ECVAM study used computer simulations of ATC testing for individual chemicals whereas Workshop 2000 participants used an evaluation that provided animal use based on fixed *in vivo* LD_{50} values and the fixed ATC doses
- the NICEATM/ECVAM study used the GHS acute toxicity categories for classification whereas the Workshop participants used the EU classification scheme which had fewer toxicity categories (i.e., accuracy of category prediction by any method would be higher with fewer categories)

ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the UDP

Workshop 2000 participants examined the effect of starting dose on animal usage in the UDP assay by making inferences from computer simulations of animal use shown in the UDP peer review BRD (ICCVAM 2000). Using the rule that requires testing to stop when four animals have been tested after the first reversal (and no other stopping rules), animal use is relatively insensitive to the slope of the dose-mortality curve. The number of animals required when the starting dose equals the true LD_{50} is approximately six. When the starting dose is 1/100 times the true LD_{50}, however, approximately nine animals are required. Thus, animal use is 30% less when the starting dose is the true LD_{50} compared to a starting dose of 1/100 times the true LD_{50} (ICCVAM 2001a, section 2.2.4, pg. 16). When the UDP testing stops based on the likelihood-ratio stopping rule, animal use depends heavily on the slope of the dose-mortality curve. Workshop 2000 participants estimated that 25-40% animals would be saved when the starting dose is equal to the true LD_{50} compared to a starting dose of 1/100 times the true LD_{50}.

At a slope of 0.5, on average 12.4 animals were predicted to be used when the starting dose is 1/100 times the true LD_{50}, but use of an average of 8.7 animals was predicted when the starting dose equals the true LD_{50} (30% reduction). At a slope of 8.3, an average of 11 animals were predicted to be used when the starting dose is 1/100 times the true LD_{50}, but an average of only six animals are used when the starting dose equals the true LD_{50} (46% reduction).

The animal savings predicted by Workshop 2000 participants were 25-40% based on starting at the true LD_{50} in comparison to starting at a dose 1/100 times the LD_{50} as the starting dose.

Depending on the regression evaluated, the average animal savings predicted in the NICEATM/ECVAM validation study at dose-response slopes of 2 and 8.3 were 6.6 - 13.0% (0.63-1.25 animals) for the 3T3 NRU and 6.7 -12.9% (0.64-1.23 animals) for the NHK NRU for the 72 reference substances tested (see **Section 10.2**). When calculated for the chemicals in each GHS toxicity category, the highest average animal savings at a dose-response slope of 2 was for chemicals in the $2000 < LD_{50} \leq 5000$ mg/kg category. Animal savings was predicted to be 22.6-26.2% for the 3T3 NRU and 21.0-26.0% for the NHK NRU, depending upon the regression used. The highest average animal savings at a dose-response slope of 8.3 was for chemicals in the $LD_{50} > 5000$ mg/kg group. Animal savings was predicted to be 26.8-32.0% for the 3T3 NRU and 23.2-30.6% for the NHK NRU, depending upon the regression used.. The major differences between the evaluation performed by the Workshop 2000 participants and the NICEATM/ECVAM study were that:

- the comparison default starting dose used for the NICEATM/ECVAM simulations was 175 mg/kg, rather than 1/100 times the true LD_{50} assumed by the Workshop 2000 participants (see **Section 10.2**).
- the NRU IC_{50} was used in various regressions of *in vitro* data against *in vivo* data to estimate starting doses. This estimation was not always close to the true LD_{50}, which was used by the Workshop 2000 participants. For example, the starting doses predicted by the NICEATM/ECVAM study for phenylthiourea were approximately 800 mg/kg by the 3T3 NRU and approximately 1250 mg/kg by the NHK NRU (see **Appendix N**). The true in vivo LD_{50} for phenylthiourea

is 3 mg/kg. Workshop 2000 participants used a best case scenario when they assumed that *in vitro* cytotoxicity predicted exactly the true LD_{50}.

9.2.2 Validation of 3T3 NRU for Phototoxicity

An NRU assay using 3T3 cells was validated by ECVAM and accepted for regulatory use to detect the phototoxic potential of test substances. The 3T3 NRU test for phototoxicity requires a 60-minute exposure to test chemicals, a 50-minute exposure to ultraviolet (UVA, 315-400 nm) light, and then removal of test chemical (Spielmann et al. 1998). After incubation for another 24 hours in fresh medium, NR medium is added and NRU is measured after a 3-hour incubation. Phototoxic potential is assessed by comparing the differences in cytotoxicity between negative control test plates that have not been exposed to UVA and test plates exposed to UVA.

Two different models, the Photoinhibition Factor (PIF) and the Mean Photo Effect (MPE), for the prediction of *in vivo* phototoxic potential were validated. The accuracy of the models for classifying the phototoxic potential of the 30 chemicals tested in nine laboratories was 88% for the PIF and 92% for the MPE when compared with *in vivo* classifications. Interlaboratory variability for classification (i.e., phototoxic vs. non-phototoxic) was assessed using a bootstrapping approach. For each chemical, classification based on a single experiment was compared to classification based on the mean PIF or mean MPE. Interlaboratory variability for classification was 0-18.8% for PIF and 0-20% for MPE.

The ECVAM Scientific Advisory Committee confirmed the scientific validity of the method in 1997 (ECVAM 1997) and its regulatory acceptance was noted in Annex V of Council Directive 67/548/EEC part B.41 on phototoxicity in 2000. An OECD test guideline, 432, was finalized in 2004 (OECD 2004). The test results from the 3T3 NRU phototoxicity test are used in a tiered testing approach to determine the phototoxic potential of test substances.

Performance of the 3T3 NRU phototoxicity assay could not be compared to the performance of the 3T3 NRU test method used in this validation study since different classification schemes were used (i.e., a two category classification for the phototoxicity vs. a six class

scheme for acute oral toxicity). Measurements of interlaboratory variability also used different techniques and were not comparable to those used for the NICEATM/ECVAM study.

NHK NRU Phototoxicity Assay

FAL participated in the European Union/European Cosmetic, Toiletry and Perfumery Association (EU/COLIPA) study (30 chemicals using NHK and 3T3 cells) and the ECVAM/COLIPA study (20 chemicals using only NHK cells) (Clothier et al. 1999). The authors showed that the NHK NRU test method could also be used to predict phototoxic potential. The accuracy for predicting *in vivo* results was similar to that of the 3T3 NRU phototoxicity test (see **Table 9-2**). The NHK NRU phototoxicity test uses the same chemical exposure duration (approximately 2 hours) as the 3T3 NRU phototoxicity test, but the duration of culture after UV exposure is 72 hours rather than 24 hours. NRU was measured after a 45-minute incubation with NR.

Table 9-2 Correct Predictions of *In Vivo* Phototoxicants by the NHK NRU Phototoxicity Assay

Study	3T3 NRU Photoxicity Test Method	NHK NRU Photoxicity Test Method
EU/COLIPA (Spielmann et al. 1998)	29/30 (97%)[1]	28/30 (93%)[1]
ECVAM/COLIPA	NA	18/20 (90%)[1] 19/20 (95%)[2]
Combined Study Data	45/45 (100%)[2]	44/45 (98%)[2]

[1]Mean Photo Effect prediction model
[2]Photoinhibition Factor prediction model
NA – not available

Although the NHK NRU phototoxicity test method achieved good concordance with *in vivo* phototoxicity, it has not been validated for regulatory use.

9.3 Studies Using *In Vitro* Cytotoxicity Test Methods with Established Performance Standards

The *Guidance Document* method of evaluating basal cytotoxicity assays for use in predicting

starting doses for acute oral toxicity assays provides the existing performance standards for the 3T3 and NRU test methods (ICCVAM 2001b).

9.3.1 *Guidance Document* (ICCVAM 2001b)

In addition to instructions for evaluating basal cytotoxicity methods for use in predicting starting doses for acute systemic toxicity assays, the *Guidance Document* provided results from testing 11 reference chemicals (ICCVAM 2001b). The 11 reference chemicals were tested with the 3T3 and NHK NRU test method protocols recommended in the *Guidance Document*. The 11 chemicals were chosen from the RC database so as to have a close fit to the RC IC_{50} –LD_{50} regression and to cover a wide range of cytotoxicity. The major differences in the *Guidance Document* protocols and the protocols used in this study are the reduced NR concentrations (from 50 μg/mL to 25 μg/mL in the 3T3 assay and to 33 μg/mL in the NHK assay), the increased chemical exposure duration for the 3T3 test method (from 24 to 48 hours), and the lack of a refeeding step for the NHK test method just prior to chemical application (see **Section 2.2** for further detail). Nevertheless, the *Guidance Document* shows the similarity of the results for the 11 chemicals in both the 3T3 and NHK NRU test methods to the RC data. The regressions were:

- $\log (LD_{50}) = 0.506 (\log IC_{50}) + 0.475$ ($R^2 = 0.985$) for the 3T3 NRU
- $\log (LD_{50}) = 0.498 (\log IC_{50}) + 0.551$ ($R^2=0.936$) for the NHK NRU, and
- $\log (LD_{50}) = 0.435 (\log IC_{50}) + 0.625$ for the RC.

The 3T3 and NHK NRU regressions were graphed on the RC regression (347 chemicals) to show that the regression lines as well as all 11 chemical data points were within the acceptance interval (± 0.5 log around the regression) of the RC regression (see **Appendix D-1**, *Guidance Document*, **Figures 3** and **4, pg.13**).

9.3.2 King and Jones (2003)

This study also tested the 11 chemicals recommend in the *Guidance Document* in the 3T3 NRU test method protocol recommended therein. The IC_{50} - LD_{50} regression obtained was comparable to the RC and to the 11 chemical regression provided in the *Guidance Document* (ICCVAM 2001b). The regression was $\log (LD_{50}) = 0.552 \log IC_{50} + 0.503$ ($R^2=0.929$)

while the RC regression was log (LD_{50}) = 0.435 log IC_{50} + 0.625. King and Jones (2003) graphed the results to show that the regression fit within the acceptance interval (± 0.5 log around the regression line) of the RC.

King and Jones (2003) also showed that a 3T3 NRU test method altered for high throughput testing by using a limited dose-response curve of three points yielded about the same IC_{50} as an eight concentration dose-response. A regression used to compare the IC_{50} values using the two different dose-response approaches yielded R^2 = 0.945.

9.3.3 A-Cute-Tox Project: Optimization and Pre-Validation of an *In Vitro* Test Strategy for Predicting Human Acute Toxicity (Clemedson 2005)

The A-Cute-Tox Project is an Integrated Project under the EU 6th framework program that started in January 2005 (termination date January 2010). The project was initiated in response to the the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) Directive and the 7th amendment of the Cosmetics Directive call for the broad replacement of animal experiments for finished products in 2003 and ingredients in 2009. Dr. Cecilia Clemedson of Expertrådet Environmental Competence Ltd, Sweden, is the scientific coordinator of the project.

The aim of the project is to develop a simple and robust *in vitro* testing strategy for prediction of human acute systemic toxicity, which could replace the animal acute toxicity tests used today for regulatory purposes. The objectives of A-Cute-Tox are:

- Compilation, critical evaluation, and generation of high quality *in vitro* and *in vivo* data for comparative analysis.
- Identifying factors (kinetics, metabolism and organ specificity) that influence the correlation between *in vitro* toxicity (concentration) and *in vivo* toxicity (dosage), and to define an algorithm that accounts for this.
- Explore innovative tools and cellular systems to identify new end-points and strategies to better anticipate animal and human toxicity.
- To design a simple, robust and reliable *in vitro* test strategy amenable for robotic testing, associated with the prediction model for acute toxicity.

654

655 The project is an extension of the NICEATM/ECVAM study and the EDIT (Evaluation-
656 guided Development of *In-vitro* Test batteries) program, which is the continuation of the
657 MEIC (Multicentre Evaluation of *In Vitro* Cytotoxicity tests) study. The partnership is made
658 up of the EDIT Consortium, ECVAM, and 35 other European toxicity research group
659 partners. The project has been divided into the following workpackages that will be
660 implemented by various configurations of research partners:

661 • WP1: Generation of a "high quality" *in vivo* database (through literature
662 searches and historical data) and establishment of a depository list of reference
663 chemicals
664 • WP2: Generation of a "high quality" *in vitro* database (includes data from the
665 NICEATM/ECVAM study, EDIT studies, and MEIC studies)
666 • WP3: Iterative amendment of the testing strategy
667 • WP4: New end-points and new cell systems
668 • WP5: Alerts and correctors in toxicity screening (I): Role of ADE
669 • WP6: Alerts and correctors in toxicity screening (II): Role of metabolism
670 • WP7: Alerts and correctors in toxicity screening (III): Role of Target organ
671 toxicity (neuro-, nephro-, hepato-toxicity)
672 • WP8: Technical optimisation of the amended test strategy
673 • WP9: Pre-validation of the test strategy

674

675 A-Cute-Tox aims to extend the NICEATM/ECVAM and MEIC studies approach toward a
676 full replacement test strategy by improving the prediction of acute toxicity using *in vitro*
677 methods and validating the testing procedure.

678

679 **9.4 Summary**

680

681 • *In vitro* NRU cytotoxicity test methods using various cell types have been
682 evaluated for correlation with rodent lethality endpoints (e.g., rat/mouse iv, ip,
683 and oral toxicity). Peloux et al. (1992) and Fautrel et al. (1993) showed good

684 correlation (r=0.877 and 0.88, respectively) of *in vitro* cytotoxicity with rodent
685 ip/iv and iv data, respectively.
686 • The 3T3 and NHK NRU test methods have been used for purposes other than
687 the prediction of starting doses for acute toxicity studies (e.g., ocular irritancy;
688 human lethal blood concentrations, *in vivo* phototoxicity).
689 • The 3T3 NRU test method has been validated (through ECVAM) for the
690 identification of *in vivo* phototoxic potential.
691 • No *in vitro* test methods have currently been validated for the prediction of
692 acute oral toxicity. Estimation of animal savings using *in vitro* cytotoxicity data
693 to estimate starting doses for the UDP did not use *in vitro* cytotoxicity data.
694 Instead, animal savings were estimated by assuming that the starting dose
695 equals the true LD_{50} (i.e., assumes cytotoxicity data can predict lethality
696 perfectly). Such theoretical predictions for animal savings for the UDP ranged
697 from 25-40% (ICCVAM 2001a) compared with the average animal savings of
698 6.6-13% predicted using computer simulation modeling of the UDP for the
699 chemicals tested in the NICEATM/ECVAM study. Halle et al. (1997) used the
700 *in vitro* cytotoxicity data in the RC to determine that animal savings of 32% can
701 be attained for the ATC method by using the LD_{50} predicted by the RC
702 regression as the starting dose. For the chemicals tested in the
703 NICEATM/ECVAM validation study, the average animal savings for the ATC,
704 determined by computer simulation modeling, was 8.0-14.8%.
705
706